数理解析研究所講究録
第1954巻 2015年 125-133

Reference prior based on a general divergence for multi-parameter non-regular models

筑波大学・数理物質科学研究科　橋本真太郎 (Shintaro Hashimoto)
(Graduate School of Pure and Applied Sciences, University of Tsukuba)
筑波大学・数理物質系　小池健一 (Ken-ichi Koike)
(Institute of Mathematics, University of Tsukuba)

1　背景

ベイズ推測において事前分布の選択問題は重要であり，今日に至るまで多くの議論がなされている．特に事前情報が少ない場合，または事前情報が全くない場合には客観事前分布 (objective prior) あるいは無情報事前分布 (non-informative prior) と呼ばれる事前分布を考える必要がある．歴史的に，一様事前分布が無情報事前分布として用いられてきたが，これはパラメータの1対1変換に対して不変性を持たないことから多くの批判があった (Robert (2001), Ghosh et al. (2006))．パラメータの1対1変換に対して不変な無情報事前分布として Jeffreys の事前分布がよく知られており，これは Fisher 情報量の平方根に比例した形をしている (Jeffreys (1961))．この事前分布は，適当な正則条件のもとで事前分布とそれに対応する (漸近) 事後分布の間の差異を Kullback-Leibler ダイバージェンスで測り，それを最大化するものとしても得られる (Bernardo (1979), Ghosh et al. (2006))．

しかし，上記の結果は密度の台が未知母数に依存するような非正則な確率分布に対しては適用できない．非正則な場合の事後分布の漸近展開については Ghosh et al. (1994) や Ghosal and Samanta (1997) で与えられており，正則な場合に1次の漸近分布が正規分布になるのに対して，非正則な場合の1次の漸近分布は指数分布になることを示している．最近，Hashimoto and Koike (2014) は Ghosal and Samanta (1997) で考えている非正則な確率分布のクラスにおいて α-ダイバージェンスに基づいた無情報事前分布を導出し，Ghosh et al. (2011) による正則な場合との比較を行った (表1)．

ただし，表1において $I(\theta)=E_\theta[\{(\partial/\partial\theta)\log f(X_1,\theta)\}^2]<\infty$ を Fisher 情報量とし，$g_3(\theta):=E_\theta[(\partial^3/\partial\theta^3)\log f(X_1,\theta)]<\infty$，$c(\theta):=E_\theta[(\partial/\partial\theta)\log f(X_1,\theta)]<\infty$ とする．

表 1 α-ダイバージェンス最大化事前分布の比較 (1-パラメータの場合)

$\alpha\,(<1)$	正則 (Ghosh et al. (2011))	非正則 (Hashimoto and Koike (2014))
$0<\alpha<1$	$\pi(\theta)\propto\sqrt{I(\theta)}$	$\pi(\theta)\propto\lvert c(\theta)\rvert$
$\alpha=0$	$\pi(\theta)\propto\sqrt{I(\theta)}$	$\pi(\theta)\propto\lvert c(\theta)\rvert$
$-1<\alpha<0$	$\pi(\theta)\propto\sqrt{I(\theta)}$	$\pi(\theta)\propto\lvert c(\theta)\rvert$
$\alpha=-1$	$\pi(\theta)\propto\exp\left[\int\frac{2g_3(\theta)-I'(\theta)}{4I(\theta)}d\theta\right]$	–
$\alpha<-1$	–	–

一方で，Bernardo (1979) や Liu et al. (2014) は適当な正則条件のもとで，正則な局外母数を持つ場合にダイバージェンスに基づいた reference prior を導いている．本稿では，Hashimoto and Koike (2014) における非正則分布に正則な局外母数を加えた場合の α-ダイバージェンスに基づく reference prior を導出することで，Hashimoto and Koike (2014) の結果を拡張する．また，位置尺度母数や片側切断指数型分布族におけるいくつかの例を与える．

2 局外母数をもつ非正則分布

2.1 設定

$X_1,\dots,X_n$ を互いに独立に，いずれも (Lebesgue 測度に関する) 密度関数 $f(x;\theta,\varphi)$ $(\theta,\varphi\in\mathbb{R}^1)$ をもつ確率変数列とする．ただし，θ は未知の非正則母数とし，φ は未知の正則母数とする．本稿では以後，θ を興味ある母数，φ を局外母数とする．また，任意の $\theta\in\Theta$ に対して，$f(x;\theta,\varphi)$ は未知母数 θ に依存する閉区間 $S(\theta):=[a_1(\theta),a_2(\theta)]$ において正値で，$S(\theta)$ の外側では値 0 をとるとし，区間の端点のうちの 1 つは θ に無関係あるいは $\pm\infty$ であってもよいとする．ただし，台の端点 $a_1(\theta),a_2(\theta)$ は θ に関して微分可能とし，密度関数 $f(x;\theta,\varphi)$ も θ,φ に関して 3 回偏微分可能であるとする．また，θ と φ の周辺事前密度をそれぞれ $\pi(\theta)$, $\pi(\varphi)$ とし，同時事前密度を $\pi(\theta,\varphi)$ と定義し，それぞれ微分可能であるとする．さらに，$\boldsymbol{X}=(X_1,\dots,X_n)$ を与えたときの (θ,φ) の事後分布の漸近展開の妥当性を保証するような条件が成り立つとする (Ghosal and Samanta (1997), Hashimoto and Koike (2014))．上記のような非正則な場合に，J. K. Ghosh らは “事後

分布の極限分布が存在するためには $S(\theta)$ が θ に関して単調増加または単調減少であることが必要条件である" ことを示している (Ghosh et al. (1994), Ghosal and Samanta (1997)). $S(\theta)$ は θ に関して単調減少であるとしても一般性を失わないため，以下では単調減少として話を進める[*1]. これらの仮定を満たす分布の簡単な例としては，次のような指数分布がある

$$f(x;\theta,\varphi)=\frac{1}{\varphi}e^{-(x-\theta)/\varphi},\quad x>\theta.$$

次に，ダイバージェンス最大化による事前分布を定義する.

定義 2.1 $m(\boldsymbol{x})$ を $\boldsymbol{X}=(X_1,\dots,X_n)$ の周辺分布とし，$D_\pi(\cdot,\cdot)$ をダイバージェンスとするとき，ダイバージェンス最大化による事前分布を次で定義する

$$\pi(\cdot)=\arg\max_{\pi(\cdot)}\int D_\pi(\pi(\cdot),\pi(\cdot|\boldsymbol{x}))m(\boldsymbol{x})d\boldsymbol{x}. \tag{1}$$

ここで，事前分布と (漸近) 事後分布の間のダイバージェンスが大きくなればなるほど，事前分布の持つ情報の量は小さくなることに注意する.

2.2 事後分布の漸近展開

$S(\theta)$ が単調減少より，θ の自然な一致推定量は $\hat{\theta}_n=\min\{a_1^{-1}(X_{(1)}),a_2^{-1}(X_{(n)})\}$ となる．ただし，$X_{(1)}=\min_{1\le i\le n}X_i$, $X_{(n)}=\max_{1\le i\le n}X_i$ とする．また，$\hat{\varphi}_n$ を次の修正尤度方程式

$$\sum_{i=1}^n\frac{\partial}{\partial\varphi}\log f(X_i;\hat{\theta}_n,\varphi)=0$$

の φ に関する解として定義すると推定量 $(\hat{\theta}_n,\hat{\varphi}_n)$ は一致性をもつことがわかる (cf. Smith (1985))．また，$u=n\sigma(\theta-\hat{\theta}_n)$, $v=\sqrt{n}b(\varphi-\hat{\varphi}_n)$ とおく．ただし，

$$\sigma:=\frac{1}{n}\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(X_i,\hat{\theta}_n,\hat{\varphi}_n),\quad b^2:=-\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\varphi^2}\log f(X_i,\hat{\theta}_n,\hat{\varphi}_n)$$

とする．このとき，$\boldsymbol{X}=(X_1,\dots,X_n)$ を与えたときの (u,v) の同時事後密度の漸近展開は次で与えられる.

[*1] 単調な台を持たない非正則分布 (例えば，一方向分布族 (Akahira and Takeuchi (1987)) については今回は除外して考える.

定理 2.1 (Ghosal (1999)) 先の仮定の下で，$\boldsymbol{X} = (X_1, \ldots, X_n)$ を与えたときの (u, v) の同時事後密度の漸近展開は次で与えられる

$$\pi(u, v|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(u - \frac{v^2}{2}\right) \left\{1 + \frac{1}{\sqrt{n}} D_1 + \frac{1}{n} D_2 + \frac{1}{n\sqrt{n}} D_3 + O(n^{-2})\right\}, \quad u < 0.$$

ただし，D_1, D_2, D_3 は Ghosal (1999) の p.962 によるものであり，$\pi(\theta, \varphi)$ とその導関数に依存する項とする．また，u の周辺事後密度の漸近展開は

$$\pi(u|\boldsymbol{x}) = e^u \Bigg[1 + \frac{1}{n}\Bigg\{ \left(\frac{\pi_{10}}{\pi_{00}\sigma} + \frac{2\left(\frac{\pi_{01}}{\pi_{00}}\right) a_{11} + 3a_{12}}{\sigma b^2} + \frac{6a_{11}a_{03}}{b^4}\right)(u+1) + \left(\frac{a_{20}}{\sigma^2} + \frac{2a_{11}^2}{b^2}\right)(u^2 - 2)\Bigg\} + O(n^{-2})\Bigg], \quad u < 0 \tag{2}$$

で与えられる．また，v の周辺事後密度の漸近展開は

$$\pi(v|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) \left[1 + \frac{1}{\sqrt{n}} \left\{\left(\frac{\pi_{01}}{\pi_{00}b} - \frac{2a_{11}}{\sigma b}\right) v + \frac{a_{03}}{b^3} v^3\right\} + O(n^{-1})\right] \tag{3}$$

となる．ただし

$$\pi_{rs} = \frac{\partial^{r+s}}{\partial\theta^r \partial\varphi^s} \pi(\hat{\theta}_n, \hat{\varphi}_n) \quad (r, s = 0, 1, \ldots),$$

$$a_{rs} = \frac{1}{n(r+s)!} \sum_{i=1}^{n} \frac{\partial^{r+s}}{\partial\theta^r \partial\varphi^s} \log f(X_i, \hat{\theta}_n, \hat{\varphi}_n) \quad (r, s = 0, 1, \ldots)$$

とする．

2.3 α-ダイバージェンス

本稿では，事前分布と (漸近) 事後分布の間の差異を測るダイバージェンスとして，次で与えられる平均化 α-ダイバージェンスを採用する

$$R^{\alpha}(\pi) := \frac{1 - \int [\int \pi^{\alpha}(\theta) \pi^{1-\alpha}(\theta|\boldsymbol{x}) d\theta] m(\boldsymbol{x}) d\boldsymbol{x}}{\alpha(1-\alpha)}, \quad \alpha \in \mathbb{R} \setminus \{0, 1\} \tag{4}$$

(cf. Amari (1982), Cichocki and Amari (2010), Ghosh et al. (2011))．α-ダイバージェンスは $\alpha \to 0$ のとき Kullback-Leibler ダイバージェンス，$\alpha \to 1$ のとき reverse Kullback-Leibler ダイバージェンス，$\alpha = 1/2$ のとき squared Hellinger ダイバージェンス，$\alpha = -1$ のとき chi-square ダイバージェンスに対応することから，重要なダイバージェンスの族をなしていることがわかる．

3 α-ダイバージェンスに基づく reference prior

今，同時事前密度は $\pi(\theta,\varphi)=\pi(\varphi|\theta)\pi(\theta)$ と書くことができるので，Bernardo (1979) より $\pi(\varphi|\theta)$ に関しては条件付き Jeffreys prior

$$\pi(\varphi|\theta)\propto\left\{E\left[\frac{\partial^2}{\partial\varphi^2}\log f(X_i;\theta,\varphi)\right]\right\}^{1/2} \tag{5}$$

を適用する．一方で，事前分布 $\pi(\theta)$ に関しては，$\pi(\theta)$ と事後分布 $\pi(\cdot|\boldsymbol{x})$ との間の差異を (4) で測り，それを最大化するものとしてダイバージェンス最大化による reference prior を定義する．ただし，ここでの $\pi(\theta)$ と $\pi(\theta|\boldsymbol{x})$ はそれぞれ $\pi(\theta,\varphi)$ と $\pi(\theta,\varphi|\boldsymbol{x})$ の周辺事前密度と周辺事後密度であることに注意する．(4) に対してベイズの定理より関係式 $f_n(\boldsymbol{x}|\theta)\pi(\theta)=\pi(\theta|\boldsymbol{x})m(\boldsymbol{x})$ が成り立つので，$R^\alpha(\pi(\theta))$ を計算すると

$$\begin{aligned}R^\alpha(\pi(\theta))&=\frac{1-\iint\pi^{\alpha+1}(\theta)\pi^{-\alpha}(\theta|\boldsymbol{x})f_n(\boldsymbol{x}|\theta)d\boldsymbol{x}d\theta}{\alpha(1-\alpha)}\\&=\frac{1-\int\pi^{\alpha+1}(\theta)E\left[\pi^{-\alpha}(\theta|\boldsymbol{x})|\theta\right]d\theta}{\alpha(1-\alpha)}\end{aligned} \tag{6}$$

となる．ただし，$f_n(\boldsymbol{x}|\theta)$ は θ を与えたときの $\boldsymbol{x}=(x_1,\ldots,x_n)$ の同時密度関数とする．(6) の積分内の期待値の計算は exact に行うことが容易ではないので，(2) 式で与えられる事後分布の漸近展開と shrinkage argument を用いて期待値の漸近近似すると次の補題を得る (shrinkage argument に関しては Ghosh (1994) や Datta and Mukarjee (2004) に詳しい).

補題 3.1 先の仮定の下で，$\alpha<1$ に対して

$$\begin{aligned}&E_\theta\left[\pi^{-\alpha}(\theta|\boldsymbol{x})\right]\\&=n^{-\alpha}\int\frac{|c(\theta,\varphi)|^{-\alpha}}{1-\alpha}\left[1+\frac{1}{n}\left\{\frac{\alpha^2}{1-\alpha}\frac{(\partial/\partial\theta)\pi(\theta)}{\pi(\theta)}\frac{1}{c(\theta,\varphi)}+S(\theta)\right\}+O(n^{-2})\right]\pi(\varphi|\theta)d\varphi\end{aligned}$$

が成り立つ．ただし，$c(\theta,\varphi)=E[(\partial/\partial\theta)\log f(X,\theta,\varphi)]$ であり，$S(\theta)$ は $\pi(\theta)$ とその導関数に無関係なパラメトリックな関数とする．

ここで，上の補題を用いて (6) を漸近近似したものを $\int\pi(\theta)d\theta=1$ の条件下で $\pi(\cdot)$ に関して最大化する問題を考えると，次の定理を得る．

定理 3.1 α-ダイバージェンス最大化による reference prior は，$\alpha\leq-1$ のときは一般に

存在せず，$-1<\alpha<1$ の[*2]ときは次のようになる

$$\pi(\theta) \propto \left(\int |c(\theta, \varphi)|^{-\alpha} \pi(\varphi|\theta) d\varphi \right)^{-1/\alpha}. \tag{7}$$

ただし，$c(\theta,\varphi)=E[(\partial/\partial\theta)\log f(X,\theta,\varphi)]$ であり，$\pi(\varphi|\theta)$ は (5) で与えられる条件付き Jeffreys prior である．また，$c(\theta,\varphi)$ は有限かつ 0 ではないとする．

証明はここでは省略するが，方針としては Hashimoto and Koike (2014) と同様．

4 例

例 4.1（位置母数分布族） θ を位置母数，φ を尺度母数とする次のような密度関数をもつ位置母数分布族を考える

$$f(x,\theta,\varphi)=\frac{1}{\varphi}f_0\left(\frac{x-\theta}{\varphi}\right).$$

ただし，f_0 は $[0,\infty]$ 上の密度関数とする．このとき，$|c(\theta,\varphi)|$ と $\lambda(\theta,\varphi)$ は次のようになる

$$|c(\theta,\varphi)|\propto\frac{1}{\varphi},\quad \lambda(\theta,\varphi)\propto\frac{1}{\varphi}.$$

θ が興味ある母数で φ が局外母数のとき，θ の reference prior は (7) 式で与えられるから

$$\pi(\theta)=\left(\int|c(\theta,\varphi)|^{-\alpha}\pi(\varphi|\theta)d\varphi\right)^{-\frac{1}{\alpha}}\propto\left(\int\varphi^{\alpha-1}d\varphi\right)^{-\frac{1}{\alpha}}=constant$$

となる．切断指数分布

$$f(x,\theta,\varphi)=\frac{1}{\varphi}e^{-\frac{x-\theta}{\varphi}}\quad(x>\theta)$$

などが，この分布族に属する．

例 4.2（片側切断指数型分布族） φ を自然母数とし，θ を切断母数とする切断指数型分布族を考える．このとき密度関数は次で与えられる (cf. Bar-Lev (1984), Akahira (2013))

$$f(x,\theta,\varphi)=\begin{cases}\frac{a(x)e^{\varphi u(x)}}{b(\theta,\varphi)} & (c<\theta\le x\le d),\\ 0 & (\text{他}).\end{cases} \tag{8}$$

[*2] $\alpha=0$ のときは，$\alpha\to 0$ の極限値として解釈する．

ただし，$-\infty \leq c < d \leq \infty$ であり，$a(\cdot)$ は正値かつ微分可能，さらに $u(\cdot)$ は区間 (θ, d) において絶対連続かつ $du(x)/dx \neq 0$ であるとする．また，

$$0 < b(\theta, \varphi) := \int_{\theta}^{d} a(x) e^{\varphi u(x)} dx < \infty$$

を規格化定数とする．この密度関数に対して $|c(\theta, \varphi)|$ と $\pi(\varphi|\theta)$ は次のようになる

$$|c(\theta, \varphi)| = \left| -\frac{\partial}{\partial \theta} \log b(\theta, \varphi) \right|, \quad \pi(\varphi|\theta) \propto \left(\frac{\partial^2}{\partial \varphi^2} \log b(\theta, \varphi) \right)^{1/2}.$$

よって，θ が興味ある母数で φ が局外母数のときの θ の reference prior は $-1 < \alpha < 1$ のとき

$$\pi(\theta) \propto \left(\int \left| -\frac{\partial}{\partial \theta} \log b(\theta, \varphi) \right|^{-\alpha} \left(\frac{\partial^2}{\partial \varphi^2} \log b(\theta, \varphi) \right)^{1/2} d\varphi \right)^{-1/\alpha}$$

の形でかくことができる．これより，片側切断指数型分布族においては，切断母数 θ に興味があるときの reference prior は $b(\theta, \varphi)$ のみに依存して決まることがわかる．以下，2 つの特別な場合について考える．

まず，(8) において，$c = -\infty$, $d = \infty$, $a(x) = e^{x^2/2}$, $u(x) = x$ とするとこれは切断正規分布

$$f(x, \theta, \varphi) = \begin{cases} \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\varphi)^2}}{\Phi(\varphi - \theta)} & (\theta \leq x \leq \infty), \\ 0 & (\text{他}) \end{cases}$$

であり，$b(\theta, \varphi) = \sqrt{2\pi} e^{\varphi^2/2} \Phi(\varphi - \theta)$ より，$|c(\theta, \varphi)|$ と $\pi(\varphi|\theta)$ は次のようになる

$$|c(\theta, \varphi)| = \frac{\phi(\varphi - \theta)}{\Phi(\varphi - \theta)}, \quad \pi(\varphi|\theta) = \left\{ 1 - \frac{(\varphi - \theta)\phi(\varphi - \theta)}{\Phi(\varphi - \theta)} - \left(\frac{\phi(\varphi - \theta)}{\Phi(\varphi - \theta)} \right)^2 \right\}^{1/2}.$$

よって，θ の reference prior は

$$\pi(\theta) \propto \left[\int \left(\frac{\phi(\varphi - \theta)}{\Phi(\varphi - \theta)} \right)^{-\alpha} \left\{ 1 - \frac{(\varphi - \theta)\phi(\varphi - \theta)}{\Phi(\varphi - \theta)} - \left(\frac{\phi(\varphi - \theta)}{\Phi(\varphi - \theta)} \right)^2 \right\}^{1/2} d\varphi \right]^{-1/\alpha}$$

となる．

次に，(8) において $c = 0$, $d = \infty$, $a(x) = 1/x$, $u(x) = -\log x$ とするとこれはパレート分布

$$f(x, \theta, \varphi) = \begin{cases} \frac{\varphi \theta^{\varphi}}{x^{\varphi+1}} & (\theta \leq x \leq \infty), \\ 0 & (\text{他}) \end{cases}$$

であり，$b(\theta,\varphi)=1/(\varphi/\theta^{\varphi})$ より，$(\partial/\partial\theta)\log b(\theta,\varphi)=-\varphi/\theta$, $(\partial^2/\partial\varphi^2)\log b(\theta,\varphi)=1/\varphi^2$ となるので，$|c(\theta,\varphi)|$ と $\pi(\varphi|\theta)$ は次のようになる

$$|c(\theta,\varphi)|=\frac{\varphi}{\theta},\quad \pi(\varphi|\theta)=\frac{1}{\varphi}.$$

よって，θ の reference prior は

$$\pi(\theta)\propto\left\{\int\left(\frac{\varphi}{\theta}\right)^{-\alpha}\frac{1}{\varphi}d\varphi\right\}^{-1/\alpha}=\frac{1}{\theta}\left\{\int\varphi^{-\alpha-1}d\varphi\right\}^{-1/\alpha}\propto\frac{1}{\theta}$$

となる．先の例における切断指数分布も切断指数型分布族に属することに注意しておく．

5 まとめと今後の展望

本稿では，Hashimoto and Koike (2014) で得られた結果を正則な局外母数をもつ場合に拡張し，興味ある母数の α-ダイバージェンス最大化による reference prior を導出した．また，例として位置母数分布族や片側切断指数型分布族における非正則母数の reference prior を計算した．今後の課題としては，本稿において興味ある母数と局外母数を入れ替えた場合の reference prior を導出することや，ダイバージェンスとして例えば Basu et al. (1998) で与えられている β-ダイバージェンスを考えたときのダイバージェンス最大化による事前分布を導出することなどである．

参考文献

[1] Akahira, M. (2013). Second order asymptotic comparison of the MLE and MCLE of a natural parameter for a truncated exponential family of distributions. *Math. Res. Note.*, No. 2013–001, Inst. of Math., Univ. of Tsukuba.

[2] Akahira, M. and Takeuchi, K. (1987). The lower bound for the variance of unbiased estimators for one-directional family of distributions. *Ann. Inst. Statist. Math.*, **39**, Part A, 593–610.

[3] Amari, S. (1982). Differential geometry of curved exponential families –curvatures and information loss. *Ann. Statist.*, **10**, 357–387.

[4] Bar-Lev, S. K. (1984). Large sample properties of the MLE and MCLE for the natural parameter of a truncated exponential family. *Ann. Inst. Statist. Math.*, **36**, Part A, 217–222.

[5] Basu, A., Harris, I. R., Hjort, N. and Jone, M. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, **85**, 549–559.

[6] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc., Ser. B*, **41**, 113–147.

[7] Cichocki, A. and Amari, S. (2010). Families of alpha- beta- and gamma-divergences: flexible and robust measures of similarities. *Entropy*, **12**, 1532–1568.

[8] Datta, G. S. and Mukarjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics.* Springer, New York.

[9] Ghosal, S. (1999). Probability matching priors for non-regular cases. *Biometrika*, **86**, 956–964.

[10] Ghosal, S. and Samanta, T. (1997). Asymptotic expansion of posterior distributions in nonregular case. *Ann. Inst. Statist. Math.*, **49** (1), 181–197.

[11] Ghosh, J. K. (1994). *Higher Order Asymptotics.* NSF–CBMS Regional Conference Series in Probability and Statistics, Vol.4, Inst. of Math. Statist., Hayward.

[12] Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis.* Springer, New York.

[13] Ghosh, J. K., Ghosal, S. and Samanta, T. (1994). Stability and convergence of posterior in non-regular problems. *Statistical Decision Theory and Related Topics V* (eds. S. S. Gupta and J. O. Berger), 183–199, Springer, New York.

[14] Ghosh, M., Mergel, V. and Liu, R. (2011). A general divergence criterion for prior selection. *Ann. Inst. Statist. Math.*, **63** (1), 49–58.

[15] Hashimoto, S. and Koike, K. (2014). Non-informative prior with maximum divergence for non-regular distributions. submitted.

[16] Jeffreys, H. (1961). *Theory of Probability, 3rd Ed.* Oxford University Press, London.

[17] Liu, R., Chakrabarti, A., Samanta, T., Ghosh, J. K. and Ghosh, M. (2014). On divergence measures leading to Jeffreys and other refrence priors. *Bayesian Analysis*, **9** (2), 331–370.

[18] Robert, C. P. (2001). *The Bayesian Choice: From Decision–Theoretic Foundations to Computational Implementation, 2nd Ed.* Springer, New York.

[19] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.